

信頼性試験の対象拡大と効率を上げる 機械学習の活用事例

日本信頼性学会元理事

伊藤 貞則

いまパソコンで行えるデータ処理能力は電子機器の発達とソフト技術の進歩で思考能力も持たせることができるようになってきている。これを信頼性試験に活用することで新たな対象の問題を解決できたり、作業の効率を大幅に上げられる分野があるため機械学習として広く活用が進んでいる。ここでは背景や基礎から身近な例を説明した後、三つの信頼性問題へ活用されている事例を紹介する。

1. 機械学習をなぜ今やらないと時代遅れか

1.1 背景

間違う、見落とす、持続作業に弱いなどの人間の短所を克服するために大量のデータの統計的処理には計算機が使われてきた。またその能力がアップしてくるといろいろな考える力を想定した人工知能(Artificial Intelligence)の模索が行われてきた。しかし高価な設備の整備や専門のプログラマーの配置が必要だったため国立機関や大学、大手企業などがそれにチャレンジできるに過ぎなかった。

一方で電子機器の成達は凄まじく、図1に示すように能力アップ、値段ダウン、また計算機は小型化した。さらにソフトウェアも便利なプログラムでオープンソフトが出現し、所謂自分のパソコンで計算できる範囲が大幅に増大した。

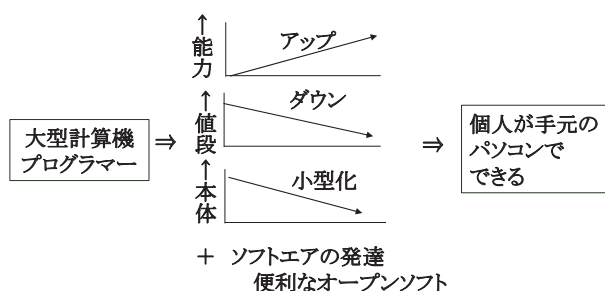


図1 計算機の発達とパソコン

その結果、計算機に求めていた理想ツールとして

- ・単に統計処理でなく予測する機能
- ・誰にでもできる机上処理
- ・新たな課題対応や開拓

などを個人が所有のパソコンでできることになってきている。すなわち高価な予算処理の必要がない、依頼処理も必要のない机上処理で、誰にでも個人の構想の下で解決できる環境が整っているのである。

1.2 機械学習(Machine Learning)とは

機械学習とは大量のデータを分析する方法の一つで、機械(計算機やパソコン)にデータを使って問題を解く方法を教えて結果を得る技術と言えよいだろう。

この機械学習は、図2に示すように概念を分けるとデータをあらかじめ指定した形(アルゴリズム)にまとめ それを使い新たな結果を予測/判定する技術とデータをいろいろな指標で分類して集中を見つけたり、表現し易いように次元削減する技術となる。前者のアルゴリズムに沿ってまとめるのは教師あり学習と言い、後者の分布や特徴を見つけるのを教師なし学習と呼んでいる。

実施方法	内容		呼び名
多くのデータをあらかじめ指定した形(アルゴリズム)にまとめ予測/判定する	連続量 直線回帰 重回帰	離散量 Classification(照合)識別 ロジスティック回帰	教師あり学習 ・目的変数 ・説明変数(特徴量)
多くのデータを種々指標で分類して集中を見つけたり表現しやすいように次元削減する	Clustering(集中)分類	次元削減	教師なし学習 ・特徴量

図2 機械学習の分類

用途としては①予測(数値データ) ②画像(波形)認識 ③音声認識 ④データ分析 ⑤自然言語処理などがある。もっとも多く実用されている画像認識では

- ・画像分類(画像判定):画像が何の画像であるかを識別
- ・異常検知(異物検出):画像の中から異なる状態を検出
- ・物体検出:画像内の対象物の場所を検出
- ・画像生成:新しい画像を生成

が出来るので観察画像の多い医療関係はもちろん工業界でも外観検査、画像のグループ化に数多く使われている。

2. 身近な機械学習の基礎

(1) 試験データに指定直線を引くと指示

本誌前テーマの[リード線の断線解析および加速試験の検討]では試験データから累積故障率1%時の時間を予測するという機械学習があった。すなわち図3に示すように発生日数と累積故障率の表をワイブル確率紙にプロットし、それらのプロットに対して最小二乗法で $y = ax + b$ のモデルの直線を得て、累積故障率1%時の時間を予測していた。

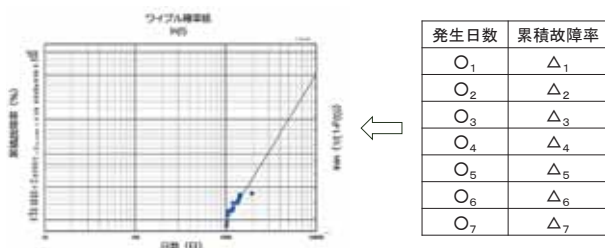


図3 表から線形回帰式を得る

これはパソコンにつぎのような作業を要求してる。

これはまず発生日数と累積故障率の表を $ax + b = y$ というモデルで表せ、すなわち線形回帰式の作成を指示し

ていることになる。統計学的に考えると発生日数 x :独立変数、累積故障率 y :従属変数として、プロットを最小2乗法に満足する直線を回帰したということになるのだが、これらの言葉は機械学習ではつぎのように表す。

$ax + b = y$: 学習前のモデル \Rightarrow アルゴリズム
発生日数 x :説明変数 累積故障率 y :目的変数
プロットをコンピュータが学習する。

またアルゴリズムからモデル式を作る場合 データの数やデータ源に偏りがあるとそのデータに対しては識別率が高いが他のデータに適用すると識別率が低いということも起きる。過学習あるいは過剰適合と言い、これを防ぐためにモデル式を作るデータを学習(トレーニング)データとし、そのほかに評価(テスト)データを準備し、汎化性能(新たなデータへの識別率)を検証するというステージが設けられている。

このように従来から行われているデータの母集団の特性把握/推定を行っている統計学と新たなデータでの対応/判断/予測を目的とする機械学習の言葉の違いを一覧にすると表1のようなになる。回帰するや推定するという数学的な言い回しから予測するや識別、判断するというように問題解決がベースにあるという言い回しになる。

表1 機械学習に使われている言葉

要素	統計学	機械学習
ex: $ax + b = y$ $y = ax + b$	モデル	学習前のモデル アルゴリズム
x	独立変数	説明変数
y	従属変数	目的変数
変数	定性変数 定量変数	質的変数 量的変数
	変数	特徴量
$ax + b = y$ a, b の値 O, Δ を求める	回帰	トレーニング 学習(訓練)
$Ox + \Delta = y$ の汎化性評価	-	テスト (検証)
確定 $Ox + \Delta = y$	回帰式	学習済モデル モデル
目的	回帰 推定	予測 識別 判断

(2) 使われているソフトは

図3の計算は500弱あるExcel関数の「LINEST: 最小二乗法による直線フィッティング用」を使用している。Excelはこのような単純な計算について非常に有能だが

幾つもの関数を呼び出したり保管を繰り返したりするのは不向きである。そのため機械学習に使われているのは便利なプログラムソフトでフリーソフトのR言語とPythonである。

☆ R言語^{[1][2]}

統計のデータ解析に特化したプログラミング言語でデータ解析を行ってグラフなどに出力する機能に優れている。プログラミングも簡単にできる。22の基本パッケージ(ライブラリ)以外600以上の拡張品が無料での使用が可能である。

☆ Python^{[3][4]}

汎用言語で活用できる範囲が広い。15の標準ライブラリがあり便利な関数を備えた幾つものモジュールがある。またpandasやNumpyなどR言語しかできなかったデータ解析用ライブラリも増えてきている。

(3) 言葉はデータベース化の際の工夫が大事

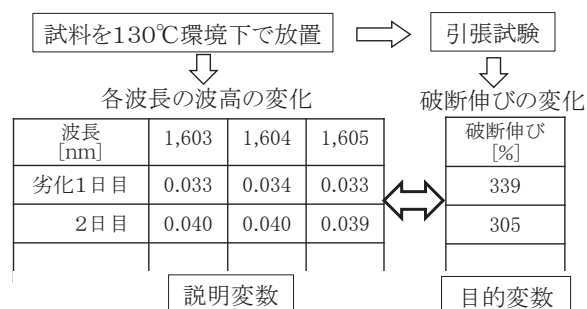
計算とともに言語データの処理には言葉が必要である。nite(製品評価技術基盤機構)では製品異常情報をもとにFMEA(Failure Mode Effect Analysis)を用いて製品事故を予測するシステムをオーソライズしているが言葉の処理に工夫が見られる^[5]。たとえば洗濯物をキーとして入力するとタオル、衣類などの情報も出て来たり、ガラガラ音がしたなどは異音で統一されたり、個別名称や擬音を共通のキーワードに統合する辞書を作り対応している。

3.活用事例

3.1 近赤外光と機械学習による材料診断^[6]

ここで紹介するのは近赤外光を透過させて得られるスペクトルから劣化状態を機械学習で予測するもので、非破壊ですぐに結果を出せるため自動車部品や建材などその場で判断が必要なものに対して有効としている。

仕組みは図4に示すようにいろいろな劣化段階を作るため130℃の環境下に一定期間暴露した高分子部品の試料から得られる近赤外線スペクトル波高値の増減と機械破断伸びの比例関係を導き出し、前者を説明変数、後者を目的変数としたアルゴリズムを作っている。



$$x_1b_1 + x_2b_2 + \dots + x_nb_n = y$$

x : 吸光度, b : 回帰係数, y : 破断伸び

y の値を高精度に推定する回帰係数を機械学習で算出

図4 モデル化の概要

近赤外スペクトルの増減測定は、まず図5(a)に示すように得られるが、ピークが密集していたり重なりがあるので特定のピークの選定や大きさの決定が難しい状態である。そのため図5(b)に示すように一般に前処理として二次微分することでピークの分離と大きさを決めている。

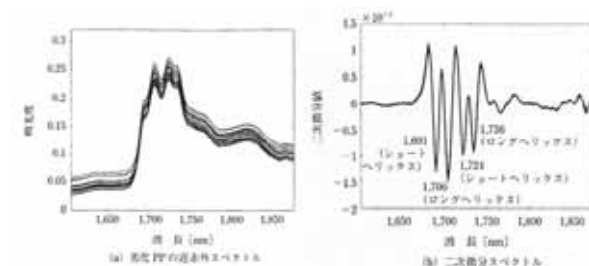


図5 劣化PPの近赤外スペクトルと二次微分スペクトル

今回の試料はPP(ポリプロピレン樹脂)である。PPには図6で示すようにアモルファス構造のSH(ショートヘリックス)構造による吸収バンドと高分子鎖が長い螺旋状のLH(ロングヘリックス)構造による吸収バンドがあり、SHが伸びる要素で、劣化が進むと結晶化度が増加してLHが多くなることで破断伸びが悪くなる。

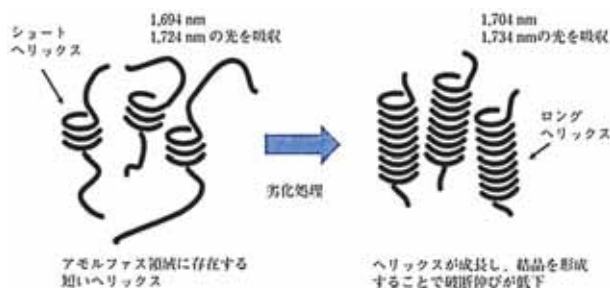


図6 回帰係数から推定される劣化メカニズム

- この図6の結晶構造と先の図5(b)のピークとの関係は
- ・ 1,706nm, 1,736nmのピークはLH構造に由来するピーク
- ・ 1,691nm, 1,724nmのピークはSH構造に由来するピーク

である。したがって図4のアルゴリズムでこれらの情報を組合せると近赤外のピークの変化からPPの劣化状態に比例した破断伸びの変化を予測できることになる。しかし測定した波長が膨大な多変量データでは、破断伸びの値を精度よく推定できる回帰係数を重回帰分析と言った一般的な統計的手法で算出することは難しい。そのためコンピュータの演算能力を駆使した機械学習の手法が使われた。

線形回帰係数の算出にPLS [Partial Least Square (部分的最小二乗法)] が用いられているためつぎの2フェーズで進められた。

- ① 主成分抽出フェーズ(波長高さと破断伸びの共分散の大きい主成分の抽出)
- ② 最小二乗法フェーズ(係数の回帰分析)

近赤外測定から求めた伸びと機械試験による破断伸びの関係を図7に示す。相関は0.98であった。

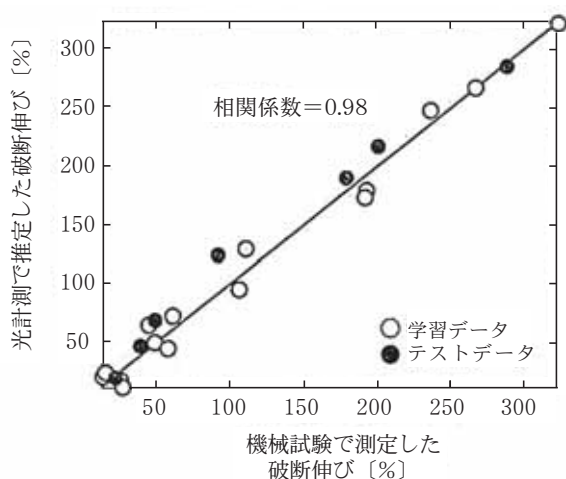


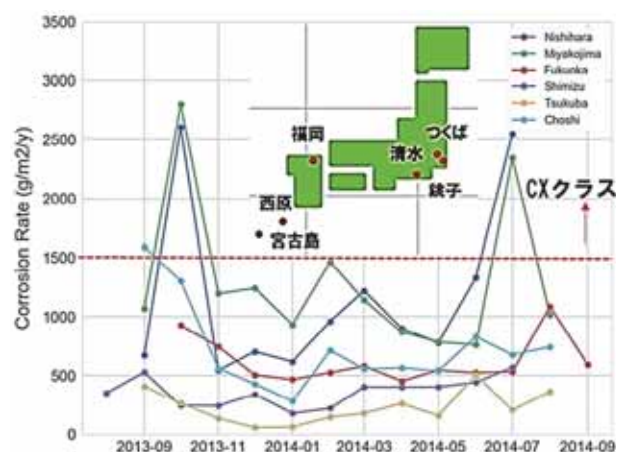
図7 実際の破断伸びと近赤外光から予測された破断伸びの関係

3.2 気象情報を用いた大気腐食量評価モデル^[7]

ここで紹介するのは腐食量解析に対して教師あり機械学習を適用した例である。説明変数は腐食因子や気象データから統計的処理を介して特徴量を抽出し、目的変

数となる実測の腐食速度データと組合せている。種々のアルゴリズムを比較検討し、腐食速度の予測器を構築している。

目的変数の腐食速度データは日本の海塩が影響する六地点(茨城県つくば市、静岡県清水市、千葉県銚子市、福岡県福岡市、沖縄県西原町、沖縄県宮古市)において炭素鋼の標準試験片による暴露試験を実施し、月ごと求めている。また説明変数は標準試験片に対する飛来海塩量、亜硫酸ガス量を測定し、それに気象データ(各測定値に最も近い気象台の温度、降水、風速など11種類:表2参照)を加えたものを1年間収集した。2013年9月から2014年8月にわたり、6地域×12ヶ月=72(内1点塩分データなしで欠損)である。図8に六地点の腐食速度データを示す。



CXクラス:ISOの最大のクラス(1500g/m²/y)以上

図8 日本各地で測定した月別腐食量

図8で西原と宮古島の2013/10月と2014/7月の高いのは、台風による海水巻き上げ雨のためである。

このように多くの説明変数を用意しているのでも、まず各説明変数と目的変数の相関関係を把握しておくことが必要である。今回の場合、両者をペアプロットし計算した相関係数 r (Pearsonの積率相関係数)の結果が表2である(p :有意確率)。この結果から月単位の計測では一因子のみによる単相関関係(例えばNaClのみの因子)では説明不可能なのが見える。相互作用を含めた各因子の影響を盛り込む上では機械学習が必要となっている。

表2 腐食量に対する各説明変数の相関関数

説明変数	r	p
NaCl	0.29	0.014
SO ₂	0.17	0.161
平均気温	0.51	0.001以下
最高気温平均	0.56	0.001以下
最低気温平均	0.44	0.001以下
降水量の合計	0.29	0.014
日照時間	-0.19	0.118
日射量	0.15	0.211
平均風速	0.67	0.001以下
最大風速	0.65	0.001以下
最大瞬間風速	0.66	0.001以下
平均湿度	0.36	0.002
最小相対湿度	0.61	0.001以下

腐食予測モデルは過学習を避けて汎化性を持たせるために4種類のアルゴリズムを検討後 良いものを選ぶ手順としている。4種類のアルゴリズムとは線形のLasso、決定木のDecision Tree、アンサンブルのRandom ForestおよびGradient Boostingである。

機械学習のソフトはPythonとそのモジュールのScikit-learnが用いられている。

アルゴリズム評価方法と結果は次のようになっている。

データ数は前述の通り71データなので63データ(90%)をTraining(学習)データとし、残り8データ(10%)をTest(評価)データとしている。図9は各モデルの適合具合を可視化/評価するために予測残差プロットしたものである。Random ForestとGradient Boostingがよく一致しているがGradient Boostingは学習データが余りにも一致しているので過学習と言える。このためRandom Forestが一番良いとしている。

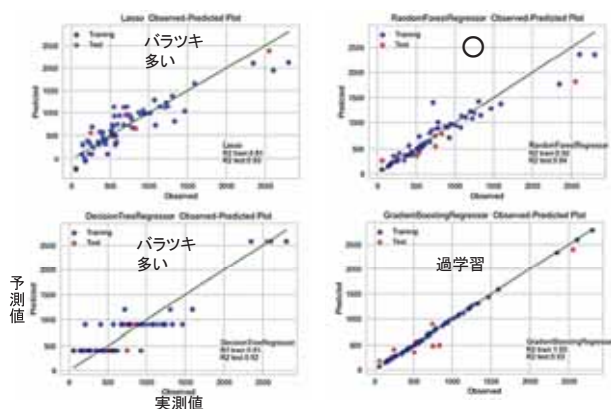


図9 Observed-Predicted Plot

3.3 ディープラーニングを用いたモータ適合効率化

ここで紹介しているのは電流波形に現れる乱れの良否判断を機械学習によって解決した例である^[8]。

電動化車両開発拡大に伴い、開発期間の大幅な短縮だけでなく熟練技術者のスキルをカバーすることが必要とされる。これを深層学習(Deep Learning)の一つであるCNN(Convolution Neural Network)を使って熟練技術者の定性的な判断を画像認識AIで代行させるとともに、実機を必要とせずCAEを用いて評価/検証が行える技術構築がされている。ソフトはPythonが使われている。

モータはPID[P(比例)、I(積分)、D(微分)]制御により電流が指令値に追従するように制御している。モータ適合とは比例ゲインKp、積分ゲインKiを変化させた際のモータ電流の指令値に対する応答遅れやオーバーシュートの大きさを判断しKpとKiの成立範囲を決定することである。Kpには指令値と実電流の差に比例した操作量を与え Kiには差の累積値に応じた操作量を与える。図10で示すように電流立ち上がり部の応答性(遅れ、オーバーシュート)と定常部乱れ(振幅など)はトレードオフになっている。この部分の波形の良否を判断するのが難しく、数少なくなっていく熟練技術者に依存している。

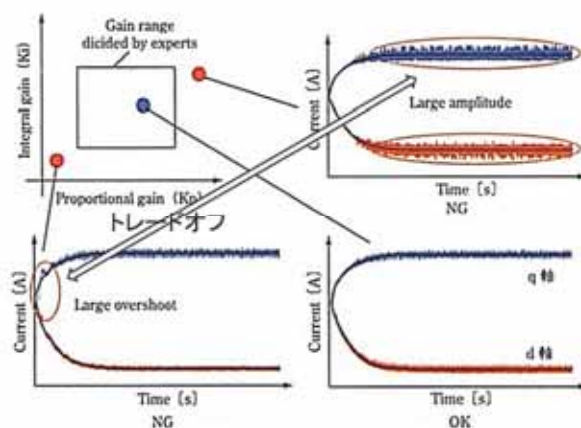


図10 電流波形と判定のポイント

このような熟練技術者に頼らざるを得ない難しい定性的な判断に画像認識AIを適用するためにはつぎのように3つの課題を解決しなければならなかった。

① 学習データの確保

深層学習では数百万の学習データが必要だが開発段階

では数十万を超える実機波形の取得が困難である。

☆ 解決策

学習データ確保のため実機ではなくCAEとしてJSOL社のJMAG-RTモデルを用いてモータ電流波形を作成した。パラメータ(電流値など)の分解能を細かくすることで定常部だけでなく 過渡部の応答性(遅れオーバーシュート)も精度よく再現できた。

② AIへのインプット方法

AIへのインプット画像の解像度に制限があるので限られた解像度で波形の特徴を盛り込んだインプット画像が必要である。モータ電流波形の場合 波形全体のバランスと高周波部分の詳細な特徴が必要となる。

☆ 解決策

解像度を有効に使うため電流波形の全体像ではなく電流波形の過渡部と定常部を分割して偏差(電流指令値-実電流)の電流波形を画像化した。

③ NN構造の選定

畳み込み層とプーリング層を積み重ねたシンプルなCNN(代表例: Alexnet)と畳み込み層を並列にしたインセプション構造を持つCNN(代表例: GoogLeNet)がある。

☆ 解決策

K_p , K_i の増減に感度のある方を選んだ。 K_p , K_i を増減させたインプット画像400枚を用意し、入力してベクトルデータ(特徴ベクトル)として取得し、K-means++法でクラスタリングを行った結果、シンプルな構造を持つCNNが K_p , K_i 方向に2分割できたが インセプション構造は K_p 方向に分割できず評価できないことがわかった。

◇ 3つの課題を解決した結果

図11に構築した波形判別用AIのTrain loss(学習誤差)とテスト誤差を示す。学習誤差が収束したことから50,000バッチ(データ:30枚/1バッチ)で学習終了と判断した。

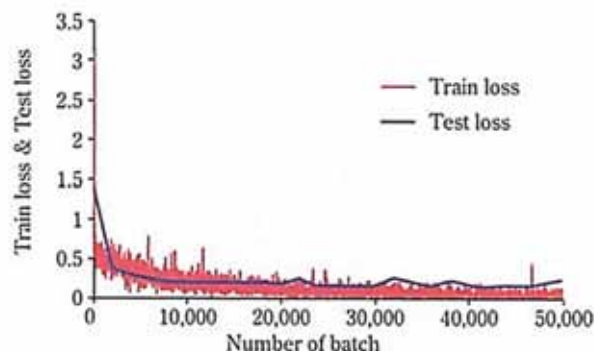


図11 学習誤差とテスト誤差

完成した波形判別用AIによって得たある車種の予測結果を図12に示す。1,000rpm~4,000rpmにおいてそれぞれ成立範囲を決定し すべての回転数で共通の領域をゲイン範囲とした。ゲイン範囲の4隅を熟練技術者が確認しAIの判定と熟練技術者の判定が一致していることを確認した。

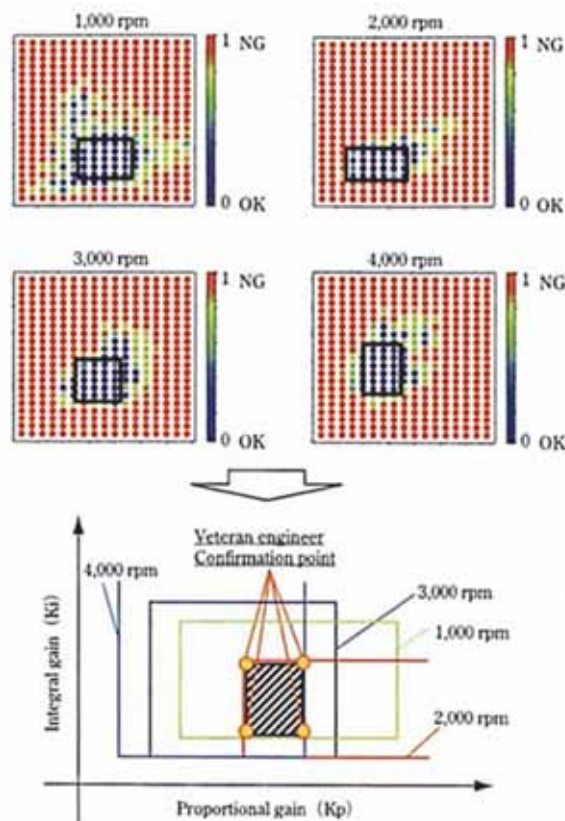


図12 予測結果

4. まとめ

機械学習の基本と活用事例を紹介した。多くの知識やデータから未来の状況を予測することの追及が行われているのがよくわかった。今後ともデータに基づいた意思

決定を求められる範囲が広がることは当然なので、今後さらに使用例が増えていくと期待している。

参考文献

- [1] 井出 剛:入門機械学習による異常検知 Rによる実践ガイド コロナ社(2015).
- [2] Brett Lante株式会社クイープ訳:Rによる機械学習 [第3版] 翔泳社(2021).
- [3] 坂本 俊之:基礎からわかるpython シーアンドアール研究所 2018.12.
- [4] Bill Lubanovic 著:斎藤 康毅 監訳,長尾 高弘 訳 入門python3 オライリー・ジャパン 2015.12.
- [5] 三浦 範大:製品事故予測システム(SAFE)の概要 nite令和元年度製品安全業務報告会.
- [6] 新澤 英之:近赤外光と機械学習による材料診断 自動車技術 Vol.75 No.3(2021年)pp66-71.
- [7] 松波 柳生 篠原 片山 須藤 服部 平口:海塩輸送シミュレーションと気象情報を用いた機械学習に基づく大気腐食量評価モデル開発と高精細腐食環境地図の作成 土木学会論文集AI(構造・地震工学)Vol.75 No.2 141-160(2019).
- [8] 寺部 俊紀:ディープラーニングを用いたモータ適合効率化 自動車技術Vol.75 No.3(2021年)pp84-89.



伊藤 貞則(いとう さだのり)

1971年 オムロン株式会社 入社
電子部品から装置までの故障メカニズム追求により信頼性向上に尽力
2006年 オムロン株式会社 退社
日本信頼性学会元理事